COMMENTARY

Comments to Jean-Claude Burgelman's article *Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story)*

George Strawn†

US National Academy of Sciences, Washington DC 20418, USA

Citation: Strawn, G.: Comments on Jean-Claude Burgelman's article *Politics and Open Science: How the European Open Science Cloud became reality (the untold story)*. Data Intelligence 3(1), 43-46 (2021). doi: 10.1162/dint_a_00075

1. PREFACE

I have heard Europeans joke about themselves saying, "America innovates, Asia imitates, and Europe hesitates." However, European Open Science Cloud (EOSC) is a case where Europe is innovating while America is hesitating. I'm a committed believer in Open Data (OD) and Open Science (OS), so I salute their leadership. And I appreciate Jean-Claude's giving us an insider's view as to how this leadership (and its associated challenges) came about. I had the good fortune of "being on the inside" in some of the projects which led to today's global Internet, so my brief remarks will focus on describing some of the salient features of the Internet development that may (or may not) apply to the EOSC development.

2. THE EOSC PROJECT AND THE INTERNET PROJECT

Jean-Claude says, "The track record of good European (high tech) ideas going nowhere because once a good idea is launched, making it a reality is obstructed by discussions on process, aggregation of particular interests, territorial fights, etc. is larger than its success record."

Jean-Claude and his colleagues had to fight political battles up front, whereas the Internet began as a political stealth project, because it was not perceived to be important at that time and only one agency funded its development. On the other hand, EOSC is widely perceived to be important and hence it is political. And as Jean-Claude said, political visibility may be the EOSC's greatest challenge. The Internet had gestation period of thirty years of US Federal agency support, but no interagency politics was involved. (Interagency politics in the US are similar to interstate politics in the EU.)

† Corresponding author: George Strawn (Email: gostrawn@gmail.com; ORCID: 0000-0003-4098-0464).

This gestation period was comprised of three sequential "middle-out" projects (middle-out meaning single agency projects). The EOSC implementors are planning for a much shorter gestation period. As such, they have compressed into one top-down EOSC project what the Internet experienced over thirty years. I'm not sure we know enough about Data Science yet to meet a very short timeline. And I hope that the EOSC effort will engage more data scientists than data users. Otherwise, EOSC will be at best an incremental improvement.

Jean-Claude also says, "Democracy and inclusiveness come with a rightful price to be paid, but it should not be at the cost of immobilism or non-delivery by watering it down." Unfortunately, that's often the challenge when the politics must precede the science.

3. THE FIRST INTERNET PROJECT: THE ARPANET

In the 1960s, APRA (now DARPA), an agency in the US Department of Defense, had cart blanche to think long-term, money to invest in long-term projects, and protection from interference by the Department of Defense. The original goal of the ARPAnet was to allow any terminal in the ARPA offices to connect to any computer, regardless of which vendor, in their contractor's offices. This was impossible at that time, because computer vendors preferred lock-in. The first version was "quick and dirty" as is often a good idea for software developments because it is not clear until after done what or how you should have implemented in the first place. The second version of the ARPAnet software was defined by the TCP/IP protocols, which have stood the test of time. The ARPAnet functionality centered on three general services: remote login to computers (today called apps), file transfer (today called the Web), and email (today called email and texting). The ARPAnet had no specific applications in mind beyond the network research required to implement these three general services.

I would hope that EOSC will also seek to provide a few general services, such as making data findable, accessible, interoperable, and reusable (FAIR). And perhaps those services will be made available by developing novel, disruptive technologies. The Internet projects created the disruptive technology of packet switching. The telecommunications industry did not recognize packet switching as a threat to their long-established mode of circuit switching until it was too late. Might a similar disruptive technology emerge from the EOSC project?

Jean-Claude says, "Policy makers do recognize disruption relatively quickly but are quite often only prepared to tackle disruption via policies that guarantee continuity of business." But, reporting what a commercial data provider told him, "If EOSC will exist, … the private players will be asked to subcontract capacity (hosting or services) as we are amongst the only ones able to provide it. So, that earns business for us." Moreover, as I will say below, those data providers will probably want to extend the EOSC to an EOC (an open cloud for more than science).

4. THE SECOND INTERNET PROJECT: THE NSFNET

In an apparently unrelated development, a high-level committee of US scientists recommend in 1983 that the Federal government provide expanded access to supercomputers for US scientists, because both Europe and Asia were doing a better job of providing such access for their scientists (another case of the US hesitating)! NSF accepted the obligation and established six supercomputer centers across the country. As a "corollary" to the supercomputing initiative, the scientists recommended that a computer network should be established to interconnect those centers with the 100 US research universities. At that time, the only functioning vender-neutral network was the ARPAnet, so decision was made to develop an "NSFnet" based on the TCP/IP protocols. Thus the NSFnet came into being as a very special purpose network. In my opinion, there would have been no support, policy or financial, for a general purpose network for higher education at that time. That support came only after the NSFnet was installed and working.

The EOSC project is very different in this regard, too. It *is* recognized as important and EU support demonstrates that. It is too late to recommend that they begin this project twenty years ago! And even if we could turn back the clock, there would not have been broad support for the EOSC in 2000.

However, looking ahead to the next innovations, The EU might want to consider establishing its own ARPA-Europe (advanced research projects agency). At least two other US agencies have done so, and so is the UK[①]. An ARPA-Europe might very well have had the foresight to begin work leading to the EOSC in 2000.

5. THE THIRD INTERNET PROJECT: THE PRIVATIZATION AND COMMERCIALIZATION OF THE NSFNET

NSF designed a three-tier architecture for the NSFnet. There was a nation-wide backbone network, regional networks, and campus networks. The regionals interconnected the research universities and the supercomputing centers in their geographical region and provided them a connection to the national backbone. Because of this architecture, the regional networks were able to allow private sector companies to also join, subject to an "acceptable use" policy stating that they could only use the network for open research. At the same time, NSF offered small awards for other colleges and universities to connect to the regional networks. The result was that by the early 1990s there were approximately 1,000 connections to the NSFnet, well beyond the original goal of 100 research universities. More surprisingly still, in 1991 the number of non-university/college connections was equal to the higher education connections. At this point, private sector entrepreneurs could see that the Internet was going to be important and began lobbying Congress to "get the government out of their way."

Over the ten years of NSFnet's existence, private sector partners had provided the fiber cables and developed its routers, so part of the private sector was fully engaged. And NSF was working hard to "get out of the way" by re-architecting the network for privatization and commercialization. So in 1995 when

① https://warontherocks.com/2020/07/a-british-advanced-research-projects-agency/

the NSFnet was retired, there were "trained" corporations ready to meet the national demand which exploded that year. And as everyone knows, by the year 2,000 the Internet was well on the way to becoming the network infrastructure for the world.

Regarding the EOSC, I suspect that private sector companies will also want to use the services provided by it, which would enhance its usefulness and provide additional support. I hope that policy will allow the universal use of EOSC in the long run, just as the Internet use became universal.

AUTHOR BIOGRAPHY

George Strawn is currently the director of the Board on Research Data and Information at the National Academies of Sciences, Engineering, and Medicine where he focuses on Open Science and FAIR data. Prior to joining the Academies, Dr. Strawn was the director of the National Coordination Office (NCO) for the Networking and Information Technology Research and Development (NITRD) Program and co-chair of the NITRD interagency committee.
ORCID: 0000-0003-4098-0464